

## SYSTEMATIC REVIEW

## OBSERVER RATING VERSUS THREE-DIMENSIONAL MOTION ANALYSIS OF LOWER EXTREMITY KINEMATICS DURING FUNCTIONAL SCREENING TESTS: A SYSTEMATIC REVIEW

Liam Maclachlan, Post grad (Dip) MSK Physiotherapy<sup>1</sup>Steven G. White, MHSc (Hons)<sup>1</sup>Duncan Reid, DHSc<sup>1</sup>

## ABSTRACT

**Background:** Functional assessments are conducted in both clinical and athletic settings in an attempt to identify those individuals who exhibit movement patterns that may increase their risk of non-contact injury. In place of highly sophisticated three-dimensional motion analysis, functional testing can be completed through observation.

**Hypothesis/purpose:** To evaluate the validity of movement observation assessments by summarizing the results of articles comparing human observation in real-time or video play-back and three-dimensional motion analysis of lower extremity kinematics during functional screening tests.

**Study Design:** Systematic review

**Methods:** A computerized systematic search was conducted through Medline, SPORTSDiscus, Scopus, Cinhal, and Cochrane health databases between February and April of 2014. Validity studies comparing human observation (real-time or video play-back) to three-dimensional motion analysis of functional tasks were selected. Only studies comprising uninjured, healthy subjects conducting lower extremity functional assessments were appropriate for review. Eligible observers were certified health practitioners or qualified members of sports and athletic training teams that conduct athlete screening. The Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2) was used to appraise the literature. Results are presented in terms of functional tasks.

**Results:** Six studies met the inclusion criteria. Across these studies, two-legged squats, single-leg squats, drop-jumps, and running and cutting manoeuvres were the functional tasks analysed. When compared to three-dimensional motion analysis, observer ratings of lower extremity kinematics, such as knee position in relation to the foot, demonstrated mixed results. Single-leg squats achieved target sensitivity values ( $\geq 80\%$ ) but not specificity values ( $\geq 50\%$ ). Drop-jump task agreement ranged from poor ( $< 50\%$ ) to excellent ( $> 80\%$ ). Two-legged squats achieved 88% sensitivity and 85% specificity. Mean underestimations as large as 19° (peak knee flexion) were found in the results of those assessing running and side-step cutting manoeuvres. Variables such as the speed of movement, the methods of rating, the profiles of participants and the experience levels of observers may have influenced the outcomes of functional testing.

**Limitations:** The small number of studies used limits generalizability. Furthermore, this review used two dimensional video-playback for the majority of observations. If the movements had been rated in real-time three dimensional video, the results may have been different.

**Conclusions:** Slower, speed controlled movements using dichotomous ratings reach target sensitivity and demonstrate higher overall levels of agreement. As a result, their utilization in functional screening is advocated.

**Level of Evidence:** 1A

**Keywords:** 3D motion analysis; functional screening; lower extremity; observation.

## CORRESPONDING AUTHOR

Liam Maclachlan, Post grad (Dip) MSK  
Physiotherapy37 Charlotte Avenue, Brooklyn,  
Wellington 6021, New Zealand

Phone: 0064 21 610 164

Fax: 0064 4 385 5359

E-mail: lmac1uk@yahoo.co.uk

<sup>1</sup> Auckland University of Technology, Auckland, New Zealand

---

## BACKGROUND

Optimal performance and avoidance of injury, both in sports and in everyday life, may depend on the quality of lower extremity movement. While anatomical variance is widely accepted, certain dynamic lower extremity movement patterns have been categorized as potential precursors to non-contact lower limb injury.<sup>1,2</sup> Excessive hip internal rotation has been linked to patellofemoral pain,<sup>3,4</sup> non-contact anterior cruciate ligament trauma<sup>5-7</sup> and iliotibial band syndrome.<sup>8</sup> Therefore, the assessment of functional tasks including squats, single-leg squats and drop-jumps have been used to evaluate an individual's injury risk and to direct the content of training programs.<sup>9</sup>

Multi-camera, three dimensional (3D) motion analysis has been found to have excellent reliability in assessing lower extremity kinematic variables<sup>5</sup> and is recognized as the gold standard in kinematic assessment.<sup>10</sup> Although highly sophisticated, the cost, resource requirement and lengthy data collection times make the use of 3D motion analysis relatively uncommon. Alternatively, clinicians commonly use visual or video observation to rate functional movements and evaluate the quality of lower extremity kinematics.<sup>11,12</sup> During this form of functional assessment, observers recognize that proximal pelvic position impacts knee loading and control<sup>13</sup> while, distally, the foot can be used as a reference marker to define knee position.<sup>14</sup>

Therapeutic interventions and training programs are strongly influenced by the results of movement analysis. Therefore, it is important that clinicians, who do not have access to 3D technology, have alternative methods of rating movement. The cost-effective nature makes observation the logical alternative choice in lower limb functional evaluation. However, when conducting functional assessments, assessors must be confident that the tasks under scrutiny are sensitive enough to rule out those at low risk of injury and specific enough to ensure that high risk individuals do not pass unnoticed. While sensitivity and specificity values highlight diagnostic power, functional assessment of tasks can only be valuable if validity is established. Being a necessary condition of validity, the value of functional screening begins by establishing reliability.<sup>15</sup>

Several authors have investigated the levels of agreement within and between observers assessing lower limb kinematics.<sup>11,16,17</sup> Experienced physiotherapists observing four lower extremity functional tasks (two-legged squat, single-leg small knee bend, lunge, and hop lunge) have demonstrated high levels of intra-rater agreement and fair-to-good inter-rater agreement.<sup>17</sup> Similarly, Poulsen and James<sup>16</sup> found that novice clinicians assessing a single-leg squat were able to track Knee Frontal Plane Projection angle (FPPA), a measure of knee alignment used to denote valgus projection. Inter-rater and intra-rater agreement also exists between clinicians evaluating unilateral squats and lateral step-down tasks, however, these levels of agreement were low.<sup>11</sup> The aforementioned results testify to observer reliability. However, only a few studies have investigated the validity of human observation using 3D motion analysis as a reference standard and, to the best of the authors' knowledge, no systematic review of this comparative literature has been conducted.

Therefore, the purpose of this systematic review is to evaluate the validity of movement observation assessments by summarizing the results of articles comparing human observation in real-time or video play-back and three-dimensional motion analysis of lower extremity kinematics during functional screening tests.

## METHODS

### Data sources and search strategy

A computerized systematic search was conducted through Medline, SPORTSDiscus, Scopus, Cinhal, and Cochrane health databases between February and April of 2014. Key search terms were gait or walk\* or "biomechanical analysis" or "functional analysis" or "movement analysis" or "motion analysis" or "kinematic analysis" or "3D motion analysis" or "2D motion analysis" or "video analysis" or observation and "lower extremity" or "lower limb" or leg. Key words from returned studies, if not already included, were incorporated into the search strategy. (Table 1). Furthermore, the reference lists and 'cited by' applications within databases were perused in an attempt to achieve an all-encompassing research yield.

### Study selection

Study selection was defined by pre-determined inclusion and exclusion criteria. Eligible studies

**Table 1.** *Key search terms)***Search strategy**

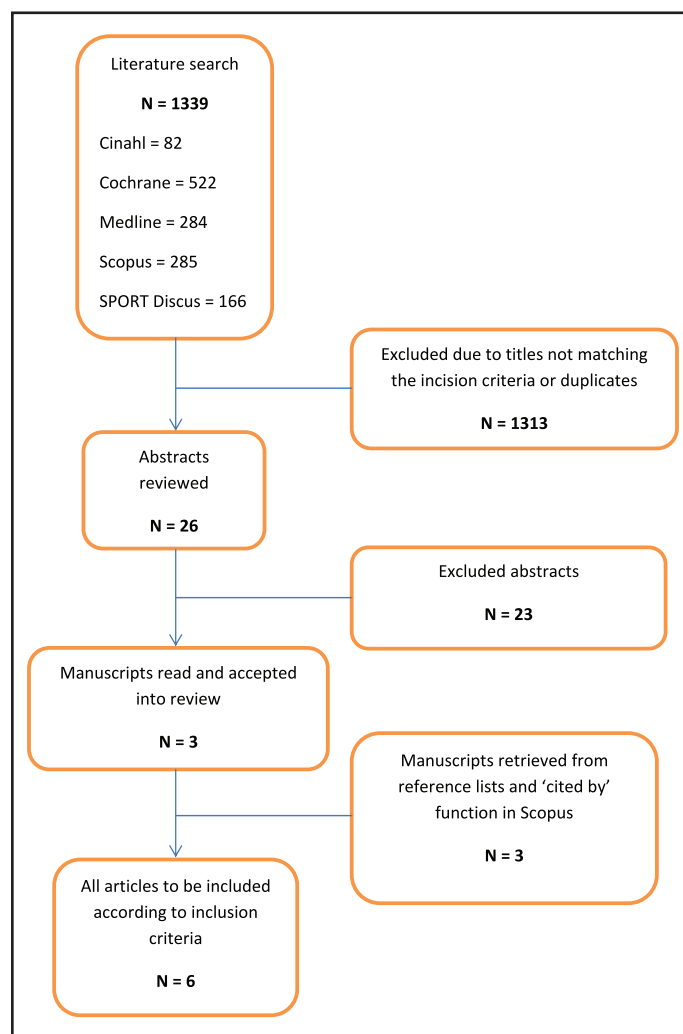
gait OR walk\* OR "biomechanical analysis" OR "functional analysis" OR "movement analysis" OR "motion analysis" OR "kinematic analysis" OR "3D motion analysis" OR "2D motion analysis" OR "video analysis" OR observation  
AND  
"lower extremity" OR "lower limb" OR leg

were those comparing human observation of real-time video play-back to 3D motion analysis of functional tasks. It was pre-determined that only lower extremity functional assessment would be appropriate and that the study samples had to comprise uninjured, healthy subjects. Eligible observers were certified health practitioners that incorporate observation in their assessments or qualified members of sports and athletic training teams that conduct athlete screening.

A stratified eligibility assessment was conducted. Initially, study titles were reviewed and titles not matching the inclusion criteria were excluded. Next, the abstracts of selected studies were read, with inappropriate studies being excluded. Finally, full manuscripts of the chosen articles were reviewed. At this stage, studies were excluded if they did not list the health/injury status of the subjects, the vocation of the observers or if it was not categorically stated that human observation and 3D motion analysis were the index test and reference standard respectively. Figure 1 displays the study selection procedure.

**Risk of bias evaluation**

To evaluate the quality of the inclusive research, the Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2) was used. The QUADAS-2 was developed through the School of Social and Community Medicine at the University of Bristol, England. The tool has been used extensively in systematic reviews aimed at the validity of diagnostic tests.<sup>18-20</sup> The QUADAS-2 consists of four key domains; patient selection, index test, reference standard, and flow and timing. All four domains utilize a series of signalling questions that assist in determining the risk of bias within the study design. The first three domains also address the study's applicability to the review. In all domains, each signalling question can be answered "yes", "no" or "unclear" and are phrased in such a

**Figure 1.** *Figure 1 – Search strategy.*

manner that "yes" indicates a low risk of bias. The creators of the QUADAS-2 indicate that if all questions within a domain are answered "yes", then the risk of bias is low. But if a signalling question is answered "no", then the potential for bias exist.<sup>21</sup>

**Data extraction**

The information extracted from the selected papers consisted of participant anthropometrics (sex, age and activity level), characteristics of observers (background, level of experience and amount of training with functional tool protocol), and details of interventions (the number and description of functional screening tools). The heterogeneity of populations, tests and outcomes used precluded a meta-analysis (Table 2). In terms of functional assessment, sensitivity and specificity, respectively, highlight the ability of a task's rating to identify movement char-

**Table 2. Study descriptions and results**

study	Objective	Subjects	Raters	Method/tests	Outcome variables	Results
Ageberg et al - 2010	To validate the observation of a single leg mini squat for assessing the position of the knee in relation to the ankle joint.	25 subjects (17 women) aged between 18-37. Recruited from the local community.	2 musculoskeletal physiotherapists with no previous experience of the test scored knee-to-foot position. Before the study both were trained for the test by an experienced examiner. A 3 <sup>rd</sup> examiner, blinded to the observation results collected the 3D data.	Single-leg mini squat. Participants were recorded using a bar for balance and looking down while bending their knee, without bending forward from the hip, until they could no longer see their toes. This mini squat was repeated 5 times at a defined speed. Opposite leg was held in slight hip flexion and approximately 80° knee flexion. Test outcome variables; knee-over-foot or knee-medial-to-foot. 2-D peak tibial, peak thigh, and peak knee varus-valgus angles (degrees), and 3-D peak hip internal-external rotation, and peak knee varus valgus angles (degrees) were calculated and used for validation of the clinical test. 3D data was collected simultaneously.	Independent t-tests to compare 2D and 3D data. A level of $p \leq 0.05$ was chosen to indicate statistical significance. The receiver operating characteristic (ROC) curve to determine the test's ability to discriminate between those with and without 2D medial knee positioning. An area under the curve close to 0.5 indicates a poor test, and a value close to 1.0 indicates a good test	10 subjects were scored knee-medial-to foot by observation. These 10 demonstrated higher tibial ( $p=0.001$ ) and thigh angles ( $p=0.001$ ) in 2D demonstrating more knee valgus ( $p<0.001$ ). In 3D, the knee internal rotation angle was also larger ( $p=0.049$ ) but valgus/varus angle was not ( $p=0.589$ ). Area under the ROC curve was 0.867 denoting a good test (SE 0.082, $p=0.002$ ).
Ekegren et al - 2009	To determine the reliability and validity of observational risk-screening in evaluating dynamic knee valgus during a drop-jump landing	40 female participants aged between 13-17. Recruited as a convenience sample from local soccer teams.	3 female physiotherapists with a mean clinical experience of $12 \pm 3$ years. All 3 were mailed a 20 minute training CD prior to testing. No mention of who collected 3D data.	A drop-jump task. Participants were recorded dropping from a 31-cm box onto an embedded force plate then immediately performing a maximum vertical jump. 9 consecutive trials were conducted with a 10 second rest between trials. Throughout, the arms were kept in a 'stop' position. Test outcome variables; high risk (knee medial to 1 <sup>st</sup> toe) or low risk (knee in line with 1 <sup>st</sup> toe). 3D data was collected simultaneously.	Physiotherapists' ratings of each participant's mean knee valgus motion was compared to 3-D motion analysis. The lead author gave each participant an expert rating of high risk or low risk. A ROC curve was constructed, linking the expert ratings with each participant's mean knee valgus motion value. Desired sensitivity $\geq 80\%$ ; desired specificity $\geq 50\%$ . A validity cut-off point of 10.83° knee valgus motion was chosen	3D motion analysis revealed 15 participants to be truly high risk and 25 truly low risk. Sensitivity targets as calculated with the ROC were met only twice: by rater 3 at time 1 (87%) and rater 1 at time 2 (87%). All raters exceeded hypothesized specificity targets (time 1 = 72%, 60% & 72%, time 2 = 64%, 68% & 72%). Although detecting some, examiners missed certain high-risk participants captured by 3D motion analysis.
Krosshaug et al - 2007	To test the accuracy and precision of researchers in estimating kinematics from video sequences that resemble situations typically leading to ACL injuries.	3 test subjects that were 22, 23 & 25 years old. Gender not stated. Place of recruitment not stated.	6 observers experienced in both ACL injury research and visual video analysis. Professional/vocational background not stated.	The participants performed trials of running and side step cutting maneuvers. 27 composite video recordings, combining different camera views. Knee flexion/extension, knee varus/valgus, knee internal/external rotation, hip flexion/extension, hip adduction/abduction, hip internal/external rotation, approach velocity, vertical velocity, cutting angle, and internal/external rotation of the foot relative to the pelvis were assessed. 35 composite videos were then created from subject 3 for a training session. After the training session, the initial 27 videos were re-assessed. 3D data was collected simultaneously.	Differences between each of the analysts' estimates and the marker-based measurements as the gold standard for each variable were measured. Paired t-tests were used to examine if the training led to significant improvements in the means (accuracy) for the differences between the estimates and the gold standard. For all analyses, an alpha level of $< .05$ was used to denote statistical significance. For the categorical variables (joint motion), a kappa-test was used to compare the agreement between the estimates and the gold standard. The strength of agreement was classified as follows: poor (value: $< .20$ ), fair (.21-.40), moderate (.41-.60), good (.61-.80), and very good (.81-1.00)	Substantial accuracy errors were found at initial testing Mean error for knee flexion was 19°, indicating a consistent underestimation. Hip angles were also underestimated systematically by an average of 7°. Both hip and knee internal rotations were underestimated by 10° and 12°, respectively. Only small overall changes in the mean error and standard deviations were seen from the pre- to the post-training tests. Overall agreement between estimated and actual direction of the joint motion was poor.
Onate et al - 2010	to assess the validity of a simple clinical jump-landing movement-assessment tool, the LESS, in identifying subjective 2-dimensional jump-landing motion analysis compared with 3-dimensional high-speed motion-analysis assessment	19 female (mean age 19.5 $\pm$ .84) soccer players conveniently sampled from a division 1 institution.	2 certified athletic trainers scored 3 trials. 1 trainer was considered an expert, with 15 years experience, and 1 was novice with less than 1 year of experience. The expert was involved in developing the test instrument and, thus, provided a 1 hour training session for the novice.	Participants were recorded performing a drop-jump task. The drop-jump was performed off a 30 cm box onto an embedded force plate that was 30 cm from the box. On landing with both feet, participants were instructed to jump as high as possible, with the initial landing being used for analysis. 3 trials were performed with a 1 minute rest in between trials. The 1 <sup>st</sup> of both raters trials was used for analysis. 3D data was collected simultaneously. The 13-point landing error scoring system was used to assess the quality of the drop-jump (see appendix 1 for criteria).	Only the expert rater's scores were used to calculate the phi coefficient correlations between the LESS scores and the 3D analysis. An alpha level of $P < .05$ was set a priori for statistical significance. Individual item analysis assessment of percent agreement was calculated and defined as poor (less than 50% agreement), moderate (51-79% agreement), or excellent (80% and above agreement)	For LESS validity, the rater achieved the following percentage agreement and Phi correlation Item scores: 1=89.5% & Phi=a, 2=21% & Phi=.118, 3=74% & Phi=-.130, 4=84% & Phi=a, 5=95% & Phi=a, 6=42% & Phi=a, 7=85% & Phi=-.233, 8=100% & Phi=a, 9=74% & Phi=a, 10=100% & Phi=a, 11=68% & Phi=-.188, 12=10% & Phi=a, 13=74% & Phi=.456.



**Table 2. Study descriptions and results (continued)**

Weeks et al - 2012	To determine kinematic predictors of perceived Single-leg squat quality for experienced and student physiotherapists and assess inter- and intra- rater reliability of single-leg squat performance rating for experienced and student physiotherapists.	22 healthy young adults (13 men and 9 women; 23.8 ± 3.1 years of age). Place of recruitment not stated.	Eight experienced post-graduate trained musculoskeletal physiotherapists and eight physiotherapy students in their penultimate year of study. No test training stated.	Participants were recorded performing 3 separate single-leg squats on each leg. Start position was opposite knee flexed to approximately 90 degrees, arms folded across their chest, and looking ahead. Participants asked to squat as far as possible without losing balance but depth was not standardized. The 2 <sup>nd</sup> squat of both legs was used for analysis. 3D data was collected simultaneously. Raters viewed each video twice at normal speed with no exceptions and rated each performance with a 10-point ordinal scale, a score of 1 representing 'very poor' and a score of 10 representing 'very good'. All raters viewed and assessed the videos again 2 weeks later	Stepwise multiple linear regression analysis was used to determine joint kinematic predictors. T-tests compared initial and repeat ratings.	Neither physiotherapist (1 <sup>st</sup> =6.4±1.3 & 2 <sup>nd</sup> =6.4±1.3) or student (1 <sup>st</sup> =6.0±1.6 & 2 <sup>nd</sup> =6.0±1.7) ratings differed significantly between the two occasions. Peak knee flexion explained 33% (physiotherapist) or 36% (student) of the variance in ratings of single-leg squat performance. Physiotherapist prediction was strengthened by 21% by adding peak hip Adduction and another 10% by adding knee mediolateral displacement. The student's prediction was strengthened by another 11% by adding knee mediolateral displacement.
Whatman et al - 2013	To investigate the ability of physiotherapists to visually rate knee and pelvic position in young athletes during lower extremity functional tests, specifically; (1) to investigate intra- and inter-rater reliability, (2) to investigate the validity of ratings and (3) to assess the influence of clinical experience and velocity of movement on rating ability.	23 (11 female) healthy young athletes (11±1 years old). All participants were part of a structured long term athlete development programme and competed in a variety of sports.	Sixty six New Zealand registered physiotherapists agreed to visually rate the video recordings. No test training stated.	Participants were recorded performing 3 lower extremity functional tests (squat, Single-leg squat and Drop jump). All participants were given standardized verbal instructions. The researcher demonstrated each test. Practice for all tests was allowed. 3D data was collected simultaneously. For all tests, physiotherapists responded "yes" or "no" to the question, "Does the patella move medial to the 2nd toe"? Additionally for the Single-leg squat, a "yes" or "no" response was recorded to the question, "Does the pelvis remain neutral in the frontal plane"? All video clips were also rated in the same manner by three musculoskeletal physiotherapists (all PhD or Masters qualified and all senior academics with an average of 15 years clinical experience). 26 of the initially selected physiotherapists repeated visual ratings 3-4 weeks later.	Validity was assessed by comparing the peak 2D and 3D kinematic data between the groups based on the consensus, expert visual ratings (To be used in this instance).	Athletes visually rated as having a patella medial to the 2nd toe alignment were likely to very likely to have increased peak 3D hip internal rotation (SKB=99%, DJ=98%, SLSKB=99%) and adduction (SKB=93%, DJ=99%, SLSKB=94%) in all tests. These athletes were also almost certain to have an increased peak FPPA in the small knee bend tests (SKB=100%, SLSKB=100%) and very likely to have an increased peak frontal plane projection angle in the Drop Jump (DJ=98%). Athletes rated as not maintaining a neutral pelvis in the frontal plane were almost certain to have increased lateral pelvic tilt as measured in 3D and 2D (both 100%).

acteristics that are considered high risk, while ensuring those that are low risk are not falsely labelled as high. When provided, sensitivity and specificity values have been reported.

## RESULTS

### Study selection

The search was conducted in Cinahl, Cochrane, Medline, Scopus and SPORT Discus and returned a total of 1339 studies. At the first selection stage, all titles were reviewed which led to 1313 studies being excluded for not matching the inclusion criteria or as duplicates. Next, all remaining abstracts were read and on the basis of the information provided, 23 additional studies did not meet the inclusion criteria and were excluded. Following manuscript readings, the final three articles were accepted into the

review. A further three articles were retrieved as a result of checking the accepted articles reference lists and the 'cited by' function in Scopus. (Figure 1).

### Methodological quality evaluation results

The risks of bias and applicability concerns of the included studies are presented in Table 3. Several authors<sup>22,23</sup> have highlighted the representational problems associated with generating an overall quality "score" of clinical trials. Therefore, instead of an overall score, this review has summarized the QUA-DAS-2 results, including high risk areas.

Risk of bias was unclear for four studies due to a lack of information relating to recruitment strategies, inclusion and exclusion criteria, and subject anthropometrics. The two remaining studies were deemed

**Table 3.** Results of the QUADAS-2 methodological quality evaluation

STUDY	RISK OF BIAS				APPLICABILITY CONCERNS		
	PATIENT SELECTION	INDEX TEST	REFERENCE STANDARD	FLOW AND TIMING	PATIENT SELECTION	INDEX TEST	REFERENCE STANDARD
Ageberg et al (2010)	?	✓	✓	✓	✓	✓	✓
Ekegren et al (2009)	X	?	?	✓	✓	✓	✓
Krosshaug et al (2007)	?	?	?	✓	?	✓	✓
Onate et al (2010)	X	?	?	✓	✓	✓	✓
Weeks et al (2012)	?	?	?	✓	✓	X	✓
Whatman et al (2013)	?	?	?	✓	✓	?	✓

Icon representations: ✓ = low risk; X = high risk; ? = unclear risk

to be high risk for using convenience samples. Risk of bias regarding index tests and reference standards was unclear for five of the studies, with only one clearly stating that both the index test and reference standard were interpreted without knowledge of the other's results. All studies achieved a low risk rating for flow and timing. One study was considered to be an unclear risk in patient selection as age was the only information provided in relation to participant background, presentation and selection.

All six studies provided clear detail regarding the application of 3D motion analysis, so were considered low risk of reference standard applicability. In terms of the index tests, five of the studies scored specific joint positions (hip and/or knee and/or ankle) in the frontal, transverse and sagittal planes. However, one study used a general scale of one to ten with one representing 'very poor' and ten representing 'very good' and, as a result was considered a high risk of index test applicability. Another study's index test applicability risk was unclear because instead of a direct comparison, participant observations were validated with expert ratings which, in turn, were validated by 3D motion analysis.

## RESULTS

### Single leg squat

Ageberg et al<sup>24</sup> and Whatman et al<sup>9</sup> asked observers to rate a single-leg squat by answering 'yes' or 'no' to whether the knee travelled medial to the 2<sup>nd</sup> toe during the movement. Observer results were then compared to the findings of 3D motion analysis. The Ageberg et al<sup>24</sup> observers were anteriorly positioned to rate real-time movement and estimated that, of

25 participants, ten had knee-medial-to-foot. When compared to 3D findings, the ten knee-medial-to-foot participants had more hip internal rotation ( $p = 0.049$ ) than those deemed to have knee-over-foot positioning ( $10.6^\circ \pm 2.1^\circ$  compared to  $4.8^\circ \pm 1.8^\circ$ ).

Observers for Whatman et al<sup>9</sup> were asked to rate an anterior view recording of a single-leg squat and they achieved the target rating sensitivity value ( $\geq 80\%$ ) but not the target mean specificity value ( $\geq 50\%$ ). These results suggest that those with knee-medial-to-foot alignment were likely to be identified, but that those without medial knee movement had a higher chance of being incorrectly classified.

Weeks et al<sup>25</sup> opted to use a ten point ordinal scale to rate the overall quality of a single-leg squat, with a score of one representing 'very poor' and a score of 10 representing 'very good'. In this instance, observers rated an anterior view recording but were not given rating guidelines for the use of the scale. Instead, they were simply asked to rate the quality of the movement, which resulted in an average score of  $6.4 \pm 1.3$ . When compared to the findings of 3D motion analysis, variance in scores was predicted by peak knee flexion (men =  $86.2^\circ \pm 13.0^\circ$ , women =  $71.5^\circ \pm 7.3^\circ$ ), peak knee medio-lateral displacement (men =  $44.8^\circ \pm 13.9^\circ$ , women =  $52.2^\circ \pm 22.7^\circ$ ) and peak hip adduction (men =  $15.5^\circ \pm 5.0^\circ$ , women =  $20.8^\circ \pm 7.1^\circ$ ). Results that suggest the accuracy of observation may depend on the joint or system of joints at which altered movement occurs.

### Drop-jump task

Whatman et al<sup>9</sup> and Ekegren et al<sup>26</sup> compared observer ratings to the results of 3D motion analysis

of a drop-jump task. Using the same knee-to-foot position rating as previously mentioned, observers in both studies rated anterior view recordings and correctly identified enough participants without knee-medial-to-foot alignment to achieve the target specificity values ( $\geq 50\%$  in both). However, the target sensitivity value ( $\geq 80\%$ ) was only met twice (87% and 87%) for Ekegren et al<sup>26</sup> and, with an overall score of 79%, was missed for Whatman et al<sup>9</sup>. These findings suggest that those with knee-medial-to-foot may go unidentified by observers of a drop-jump.

Onate et al<sup>27</sup> used the 13-point LESS to rate a drop-jump. Medio-lateral right side and anterior view recordings were provided for observer rating. Observer agreement with 3D motion analysis was found to be excellent (above 80%) for six of the points, moderate (51%-79%) for four of the points and poor (under 50%) for three points. Of all the points, best agreement (100%) was found when grading 'toes greater than 30° internal rotation in terms of foot position at initial contact', and 'stance width greater than shoulder width'. The three points that resulted in poor agreement were knee flexion at initial contact (21%), initial foot contact (42%) and lateral trunk flexion at initial contact (10%). Taken collectively, these findings suggest that the precision of observation may be dependent on body region under analysis.

### Other functional tasks

Whatman et al<sup>9</sup> compared observer assessment to 3D motion analysis findings of a two-legged squat with arms at the side. When asked to determine whether the knee travelled medial-to-foot from anterior view recordings, 88% sensitivity and 85% specificity scores were achieved; both above the author's target values of  $\geq 80\%$  and  $\geq 50\%$  respectively. While the majority of physiotherapists (1<sup>st</sup> quartile  $\geq 80\%$  &  $\geq 50\%$ ) achieved the target value, those with five years more experience improved rating accuracy (diagnostic odds ratio > 2).

Krosshaug et al<sup>28</sup> found substantial accuracy errors between observer ratings of video play-back and 3D findings of a running and side-step cutting manoeuvre. With one camera at postero-lateral left side, one at the right side (mid-stance) and one anteriorly

placed, observers used three different views to rate the movement. Underestimations were seen in the mean error for knee flexion (-19°), hip flexion (at an average of 7°) and knee and hip internal rotation (by 10° and 12° respectively). Multivariate regression analysis showed that knee flexion estimate errors were significantly less when a side camera was present ( $p=0.02$ ), when the right leg (the one closest to the camera) was analyzed ( $p= <0.001$ ) and when knee flexion was lower than 30° ( $p= <0.001$ ). Similarly, hip adduction/abduction estimate error was significantly lower when a front camera was included ( $p=0.017$ ). In this instance, a significant relationship was observed between the mean estimate error and the true joint angle for 3-D valgus ( $p= <0.001$ ) and hip flexion ( $p= <0.001$ ). Overall, these results demonstrate that observation precision is also influenced by range of motion and the plane in which movement is observed.

### DISCUSSION

The studies in this review explored levels of agreement between human visual and video observation and 3D motion analysis when assessing a range of functional tasks. Clinically acceptable results, in terms of the accuracy of observer ratings, were achieved when slower, speed-controlled movements such as a single-leg squat<sup>24,25,9</sup> and a two-legged squat<sup>9</sup> were rated. Conversely, lower levels of agreement were evident when faster, more explosive movements such as a drop-jump<sup>26,27,9</sup> or running and cutting manoeuvres<sup>28</sup> were assessed. While some tasks may have been easier to rate than others, several additional factors deserve consideration. Varying task instruction, for similar tasks, may have influenced rating accuracy. Furthermore, certain characteristics of the observers used and the chosen samples may have affected final sensitivity and specificity values.

The accurate rating of a functional task will be strongly influenced by how far the movement under scrutiny varies from what is considered normal. Movement that borders on what is considered to be normal may be harder to rate whereas variations, from normal, of a larger magnitude, become more obvious and easier to rate. It may also be fair to assume that, as the difficulty level of a task increases, so too does the likelihood of movement

---

beyond normal parameters. Therefore, rating accuracy, in each study may have been affected by how tasks were being performed.

Different methods were apparent for the squat tasks. Ageberg et al<sup>24</sup> encouraged participants to achieve approximately 50 degrees knee flexion, while Whatman et al<sup>9</sup> and Weeks et al<sup>25</sup> encouraged participants to squat as far as possible while maintaining control and whole-foot ground contact. Considering that, as knee flexion increases, so does hip flexion, and that increases in hip flexion compromise hip abduction torque as the moment arm of gluteus medius decreases,<sup>29</sup> those with more knee flexion may have demonstrated a level of contralateral pelvic drop that was easier to observe.

Upper extremity support used by Ageberg et al,<sup>24</sup> during single-leg squat may have reduced task difficulty and therefore, rating accuracy. As Dingenen, Malfait, Vanrenterghem, Verschueren and Staes<sup>30</sup> highlight, knee loading during a single-leg squat is the result of whole body loading. Therefore, engaging the upper extremities may have aided trunk control, reducing the magnitude of aberrant movement and making it harder for observers to detect variance.

Higher drop heights, when drop-jumping, have been found to significantly increase jump height (JH) and relative peak eccentric force (RPEF).<sup>31</sup> However, Barr and Nolte<sup>31</sup> failed to notice significant changes between dropping heights of 24 cm (JH =  $0.38 \pm 0.05$ ; RPEF =  $4.15 \pm 0.91$ ) and 36 cm (JH =  $0.37 \pm 0.04$ ; RPEF =  $4.36 \pm 0.72$ ). Therefore, the different box heights used across the studies, 31 cm,<sup>26</sup> 30 cm<sup>27</sup> and 25 cm,<sup>9</sup> is unlikely to have influenced the control of movement and task rating accuracy.

The sensitivity and specificity of any test that requires human rating may be strongly influenced by the skill-level and experience of the person conducting the task. Novice raters using the Functional Movement Screen™ (a battery of seven tests that categorize fundamental movement) have previously demonstrated substantial-to-excellent agreement with expert findings.<sup>34</sup> However, when assessing the quality of the movement pattern of a range of lower extremity functional movements, experienced physiotherapists have been found to have a higher level of intra-rater agreement (87% [CI 76-94]) than

inexperienced (82% [CI 71-94]) and novice counterparts (80% [CI 63-92]).<sup>17</sup> The variance associated with lower levels of agreement, whether intra or inter-rater, may increase the likelihood of poor movement being rated as good and vice versa.

In this review the experience levels of observers ranged from second-year students through to expert raters with 15 years of experience and post-grad qualifications. Interestingly, of the studies that compared novice to expert ratings,<sup>9,25,27</sup> only Whatman et al<sup>9</sup> noticed a significant difference in drop-jump ratings, with experienced physiotherapists achieving a substantial-to-excellent agreement (percentage agreement 82-90%, first order agreement co-efficient 0.65-0.81) compared to the fair-to-substantial agreement achieved by the less experienced group (PA: 76-86%, AC1: 0.56-0.78).

Previous investigators have used expert rating<sup>35</sup> and 3D motion analysis<sup>36</sup> to demonstrate that individuals with greater hip abduction, knee flexion and knee extension torques are less likely to demonstrate aberrant hip and/or knee biomechanics. Therefore, the baseline activity and strength levels of the participants involved in the studies may have determined the ease with which observers could define lower extremity kinematics and, thus, the overall sensitivity and specificity values. In this review, the three studies rating a drop-jump all recruited active individuals,<sup>9,26,27</sup> From this group, both Whatman et al<sup>9</sup> and Ekegren et al<sup>26</sup> reported acceptable levels of sensitivity and specificity for the task. Of the three studies rating a single-leg squat, Whatman et al<sup>9</sup> (who used an athletic population) and Ageberg et al<sup>24</sup> (who used a mixed cohort of recreationally physically active and not physically active) reported activity levels of their participants. Weeks et al<sup>25</sup> did not provide this detail. As only Whatman et al<sup>9</sup> offer single-leg squat sensitivity and specificity values, it remains hypothetical whether or not such values would be similar in the studies using mixed populations.

When selecting a task for functional assessment, consideration should be given to the functional relationship of the task movement and the underlying activity the person partakes in.<sup>38</sup> Running and cutting manoeuvres were, in this review, arguably, the movements that come closest to resembling sporting



activity, however, they were also found to demonstrate the worst rater agreement with 3D data. With significant underestimations seen in a variety of hip and knee movements, it is more likely that, when assessing running-and-cutting, poor movement would be rated as good and that sensitivity values would be unacceptable.

Whatman, Hing and Hume<sup>39</sup> used 3D motion analysis to explore the relationship between five forms of squatting and jogging. In doing so they found a strong correlation between peak hip, knee and ankle kinematics during the tests and jogging ( $r = 0.53-0.93$ ) and concluded that squat tests may help physiotherapists determine dynamic lower extremity alignment and risk of injury within jogging cohorts. These findings, suggest that various forms of squatting may offer a strong alternative to sports specific movement when it comes to functional testing.

This review has several limitations that must be considered. The heterogeneity of populations, tasks and outcomes used precluded a quantitative review such as a meta-analysis. Furthermore, the small number of studies identified limits the generalizability of the results. This review focussed on lower extremity kinematics during functional screening. However, in practical terms, this compartmentalised approach may be over-simplistic and lead to misinterpretations of injury risk. Those assessing functional tasks must also understand the influence of and evaluate trunk positioning and whole-body movement. Finally, it is important to recognise that the majority of the studies in this review used 2D video-playback for observation. If the movements had been rated live, in real-time 3D, the results may have been different.

## CONCLUSION

Lower extremity functional 3D motion analysis is commonly used to identify movement patterns that increase the likelihood on non-contact injuries. However, the process can be expensive, resource-dependant, and time-consuming. A growing body of evidence exists, implying that human observation of functional movement may be a reliable alternative to high technology. Although the reliability of observation has been established, only a few studies have attempted to validate human observation

by establishing levels of agreement with the gold standard; 3D motion analysis. From the existing literature, the results of this systematic review demonstrate that the validity of human observation depends on several performance and rating factors. Assessing slower, speed-controlled movements such as a two-legged squat or a single leg squat produced acceptable levels of agreement with 3D. However, agreement was poorer with faster, explosive movements such as drop-jumps, running and cutting manoeuvres. Although, the pelvis, hip, knee and foot positions are often used to classify the quality of functional movement, trunk and overall body positioning impact heavily on lower extremity function and, thus, demand attention. These conclusions need to be considered in light of the risk of bias associated with the included studies. Further high quality studies are needed before a definitive statement on the accuracy of visual assessment of faulty movement patterns of the lower limb can be made.

## REFERENCES

1. Hewett TE, Myer GD, Ford KR, et al. Biomechanical measures of neuromuscular control and valgus loading of the knee predict anterior cruciate ligament injury risk in female athletes: A prospective study. *Am J Sports Med.* 2005;33(4):492-501.
2. Willson JD & Davis IS. Lower extremity mechanics of females with and without patellofemoral pain across activities with progressively greater task demands. *Clin Biomech.* 2008;23(2):203-211.
3. Powers CM. The influence of altered lower-extremity kinematics on patellofemoral joint dysfunction: a theoretical perspective. *J Orthop Sport Phys Ther.* 2003;33(11):639-646.
4. Souza RB & Powers CM. Differences in hip kinematics, muscle strength, and muscle activation between subjects with and without patellofemoral pain. *J Orthop Sport Phys Ther.* 2009;39(1):12-19.
5. Ford KR, Myer GD & Hewett TE. Reliability of Landing 3D Motion Analysis: Implications for Longitudinal Analyses. *Med Sci Sports Exerc* 2007;39(11):2021-2028.
6. Griffin LY, Albohm MJ, Arendt EA, et al. Understanding and preventing noncontact anterior cruciate ligament injuries: a review of the Hunt Valley II Meeting, January 2005. *Am J Sports Med.* 2006;34(9):1512-1532.
7. Hewett TE, Myer GD & Ford KR. Anterior cruciate ligament injuries in female athletes. Part 1,

- mechanisms and risk factors. *Am J Sports Med.* 2006;34(2):299-311.
8. Reiman MP, Bolgla LA & Lorenz D. Hip function's influence on knee dysfunction: a proximal link to a distal problem. *J Sport Rehabil.* 2009;18(1):33-46.
  9. Whatman C, Hume P, & Hing W. The reliability and validity of physiotherapist visual rating of dynamic pelvis and knee alignment in young athletes. *Phys Ther Sport.* 2013b;14(3):168-174.
  10. Miller A & Callister R. Reliable lower limb musculoskeletal profiling using easily operated, portable equipment. *Phys Ther Sport.* 2009;10(1):30-37.
  11. Chmielewski TL, Hodges MJ, Horodyski M, et al. Investigation of clinician agreement in evaluating movement quality during unilateral lower extremity functional tasks: a comparison of 2 rating methods. *J Orthop Sports Phys Ther.* 2007;37(3):122-129.
  12. Drouin JM & Riemann BL. Lower Extremity Functional Performance Testing, Part 2. *Athl Ther Today.* 2004;9(3):49-51.
  13. Leetun DT, Ireland ML, Willson JD, et al. Core stability measures as risk factors for lower extremity injury in athletes. *Med Sci Sports Exerc.* 2004;36(6):926-934.
  14. Noyes FR, Barber-Westin SD, Fleckenstein C, et al. The drop-jump screening test: difference in lower limb control by gender and effect of neuromuscular training in female athletes. *Am J Sports Med.* 2005;33(2):197-207.
  15. Donelson R. Evidence-based low back pain classification: improving care at its foundation. *Eura Medicophys.* 2004;40(1):37-44.
  16. Poulsen DR & James CR. Concurrent validity and reliability of clinical evaluation of the single leg squat. *Physiother Theory Pract.* 2011;27(8):586-594.
  17. Whatman C, Hing W, & Hume P. Physiotherapist agreement when visually rating movement quality during lower extremity functional screening tests. *Phys Ther Sport.* 2012;13(2):87-96.
  18. Hutting N, Scholten-Peeters GGM, Vijverman V, et al. Diagnostic Accuracy of upper cervical spine instability tests: A systematic review. *Phys Ther.* 2013;93(12):1686-1695.
  19. Kopkow C, Freiberg A, Kirschner S, et al. Physical Examination Tests for the diagnosis of posterior cruciate ligament rupture: A systematic review. *J Orthop Sports Phys Ther.* 2013;43(11):804-813.
  20. Wright AA, Wassinger CA, Frank M, et al. Diagnostic accuracy of scapular physical examination tests for shoulder disorders: a systematic review. *Br J Sports Med.* 2013;47(14):886-892.
  21. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med.* 2011;155(8):529-536.
  22. Jüni P, Witschi A, Bloch R, & Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. *JAMA.* 1999;282(11):1054-1060.
  23. Whiting P, Harbord R, & Kleijnen J. No role for quality scores in systematic reviews of diagnostic accuracy studies. *BMC Med Res Methodol.* 2005;5:19.
  24. Ageberg E, Bennell KL, Hunt MA, et al. Validity and inter-rater reliability of medio-lateral knee motion observed during a single-limb mini squat. *BMC Musculoskelet Disord.* 2010;11:265-273.
  25. Weeks BK, Carty CP, & Horan SA. Kinematic predictors of single-leg squat performance: A comparison of experienced physiotherapists and student physiotherapists. *BMC Musculoskelet Disord.* 2012;13:207-214.
  26. Ekegren CL, Miller WC, Celebrini RG, et al. Reliability and validity of observational risk screening in evaluating dynamic knee valgus. *J Orthop Sports Phys Ther.* 2009;39(9):665-674.
  27. Onate J, Cortes N, Welch C, et al. Expert versus novice interrater reliability and criterion validity of the landing error scoring system. *J Sport Rehabil.* 2010;19(1):41-56.
  28. Krosshaug T, Nakamae A, Boden B, et al. Estimating 3D joint kinematics from video sequences of running and cutting maneuvers—assessing the accuracy of simple visual inspection. *Gait Posture.* 2007;26(3):378-385.
  29. Olson TJ, Chebny C, Willson JD, et al. Comparison of 2D and 3D kinematic changes during a single leg step down following neuromuscular training. *Phys Ther Sport.* 2011;12(2):93-99.
  30. Dingenen B, Malfait B, Vanrenterghem J, et al. The reliability and validity of the measurement of lateral trunk motion in two-dimensional video analysis during unipodal functional screening tests in elite female athletes. *Phys Ther Sport.* 2014; 15(2):117-123.
  31. Barr MJ, & Nolte VW. The importance of maximal leg strength for female athletes when performing drop jumps. *J Strength Cond Res.* 2014;28(2):373-380.
  32. Kennedy MD, Burrows L, Parent E. Intrarater and Interrater Reliability of the Single-Leg Squat Test. *Athl Ther Today.* 2010;15(6):32-36.
  33. Cortes N & Onate J. Clinical Assessment of Drop-Jump Landing for Determination of Risk for Knee Injury. *Int J Athl Ther Train.* 2013;18(3):10-13.
  34. Minick KI, Kiesel KB, Burton L, et al. Interrater reliability of the functional movement screen. *J Strength Cond Res.* 2010;24(2):479-486.

- 
35. Crossley KM, Zhang WJ, Schache AG, et al. Performance on the Single-Leg Squat Task Indicates Hip Abductor Muscle Function. *Am J Sports Med.* 2011;39(4):866-873.
  36. Claiborne TL, Armstrong CW, Gandhi V, et al. Relationship between hip and knee strength and knee valgus during a single leg squat. *J Appl Biomech.* 2006;22(1):41-50.
  37. Haines TL, McBride JM, Triplett NT, et al. A comparison of men's and women's strength to body mass ratio and varus/valgus knee angle during jump landings. *J Sports Sci.* 2011;29(13):1435-1442.
  38. Schneiders AG, Davidsson Å, Hrman E, et al. Functional movement screen normative values in a young, active population. *Int J Sports Phys Ther.* 2011;6(2):75-82.
  39. Whatman C, Hing W, & Hume P. Kinematics during lower extremity functional screening tests–Are they reliable and related to jogging? *Phys Ther Sport.* 2011;12(1):22-29.
  40. Ortiz A, & Micheo W. Biomechanical Evaluation of the Athlete's Knee: From Basic Science to Clinical Application. *PM & R.* 2011;3(4):365-371.